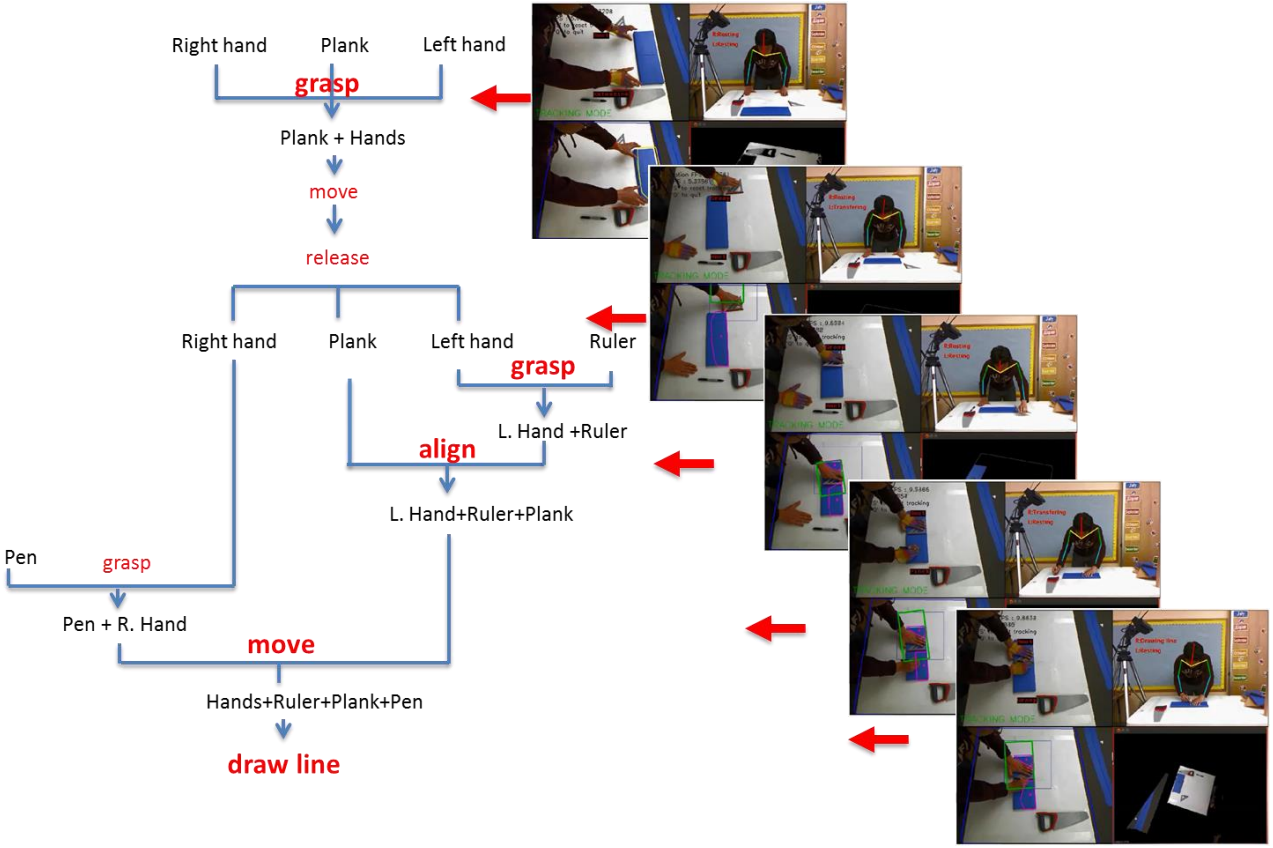
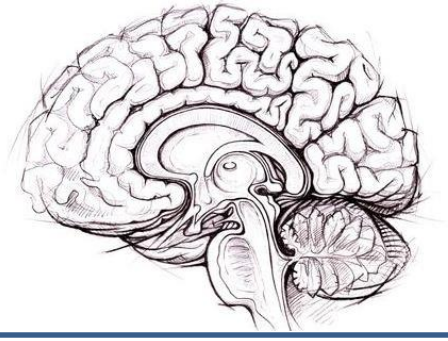
Learning for Event Cognition

Cornelia Fermüller
UMIACS, University of Maryland

The multiple scales of Event Perception



Assimilate understanding of cortical function



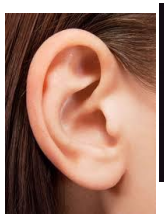
Create robust systems and devices

Smart hearing aid

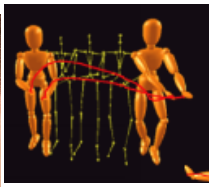
Cognitive robots

Gain an understanding of Learning and Expertise for Systems with Perception and Action

Vision



Audition



Action

The Two Aims of the Project

Theoretical Framework

Representations
Dynamics
Cognition

Neural Models of Cortical Processing

Robust Perception and Action

Noise, mixtures, contexts

Experimental Approaches

Neuromorphic Hardware

Human

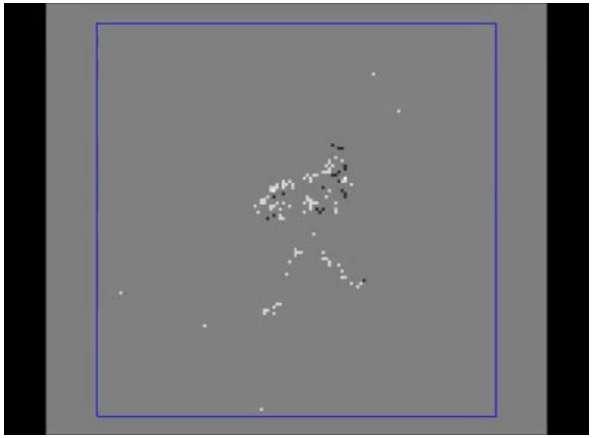
MEG, EEG, Psychoacoustics

Animal

Behavior
Neurophysiology

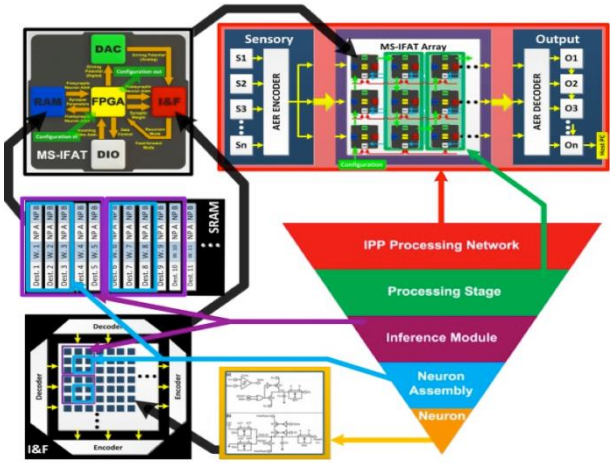
Neuromorphic hardware

Asynchronous Dynamic Vision Sensor



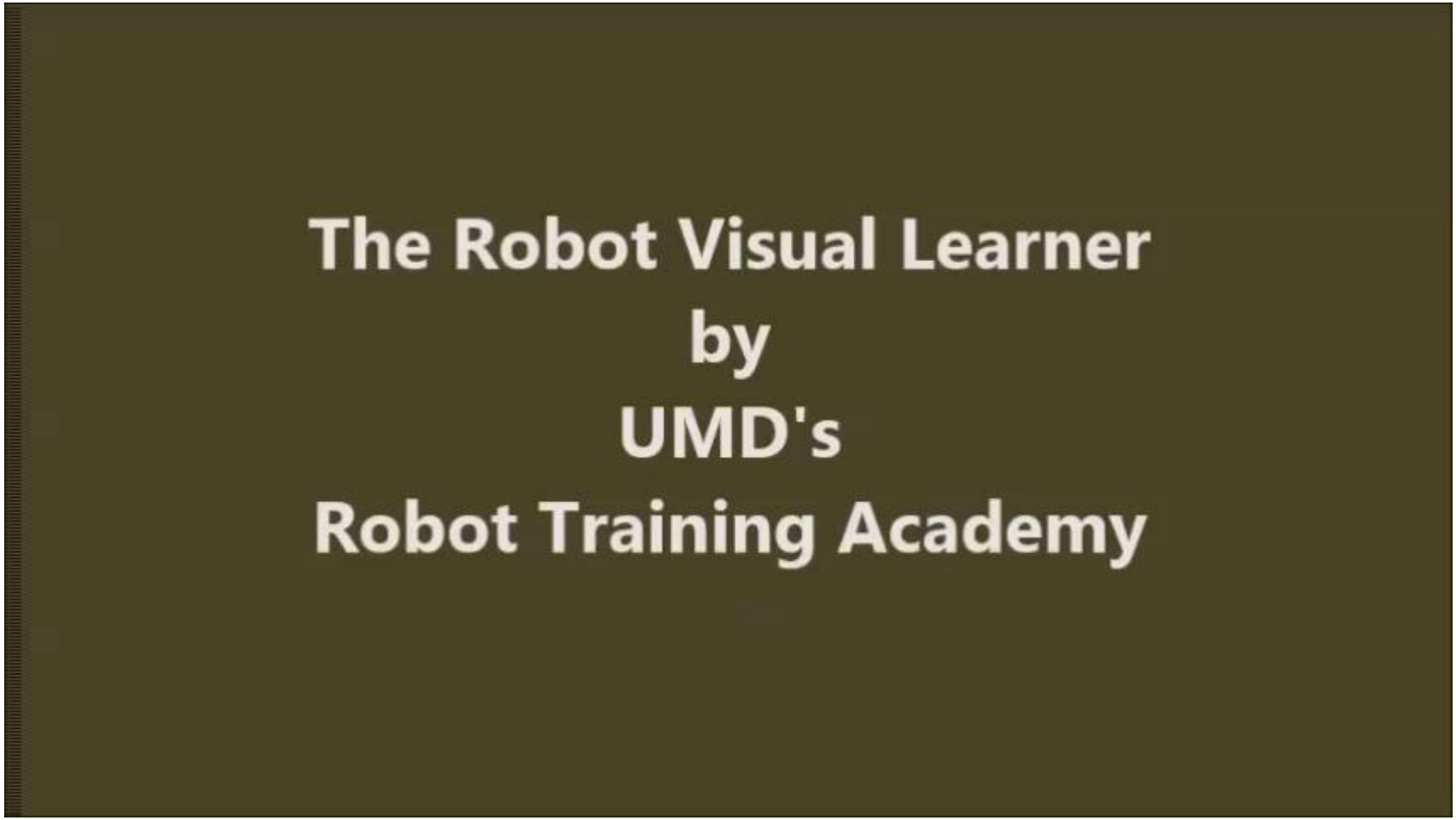
Biological inspired large-scale circuits:

SpiNNaker
TrueNorth
Neurogrid



Implementation of all the hierarchical levels of the Image Processing Pipeline.

Video



Four Cognitive Primitives

1. Adaptive and contextual integration

2. Saliency & attentional modulation

3. Binding, segmentation, temporal coherence

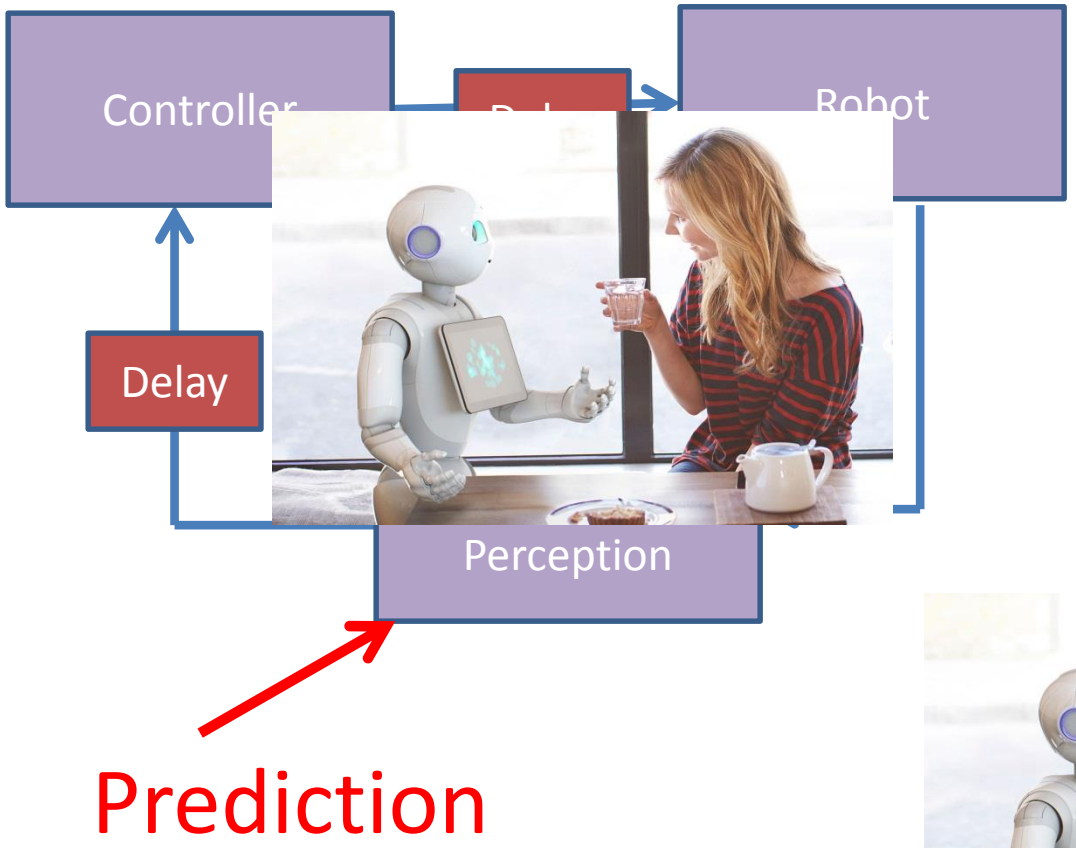
4. Category formation

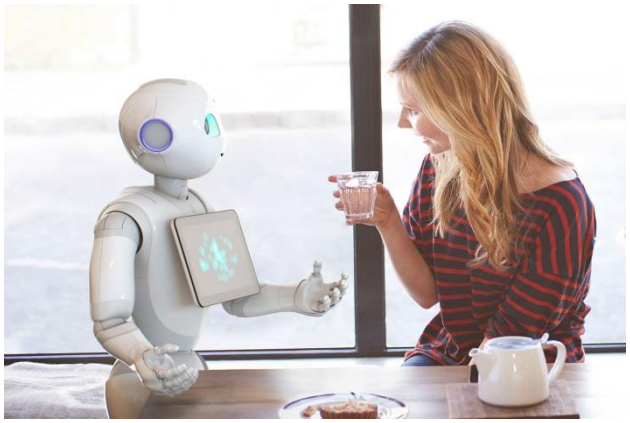
NSF: "**SL-CN: Cortical Architectures for Robust Adaptive Perception and Action**."
C. Fermüller (UMD) , S. Shamma, T. Horiuchi (UMD), R. Etienne-Cummings, A. Andreou.

Motivation: Human Robot Collaboration

Telluride Project led by C. Fermüller, A. Andreou, and M. Pfeiffer.
“Prediction of Manipulation Actions.”

Action Perception Loop





Humans predict others to be proactive

Multi-Sensor Recording

6 objects each of 5 actions, 5 actors

Data Preparation for learning: Label for each action : start, end, contact.

Object and action pairs

Object	Actions
cup	drink, pound, shake, move, pour
stone	pound, move, play, grind, carve
sponge	squeeze, flip, wash, wipe, scratch
spoon	scoop, stir, hit, eat, sprinkle
knife	cut, chop, poke a hole, peel, spread
fork	eat, poke a hole, pick, scratch, whisk

Recurrent Neural Network (RNN)
The Long Short-term Memory Model

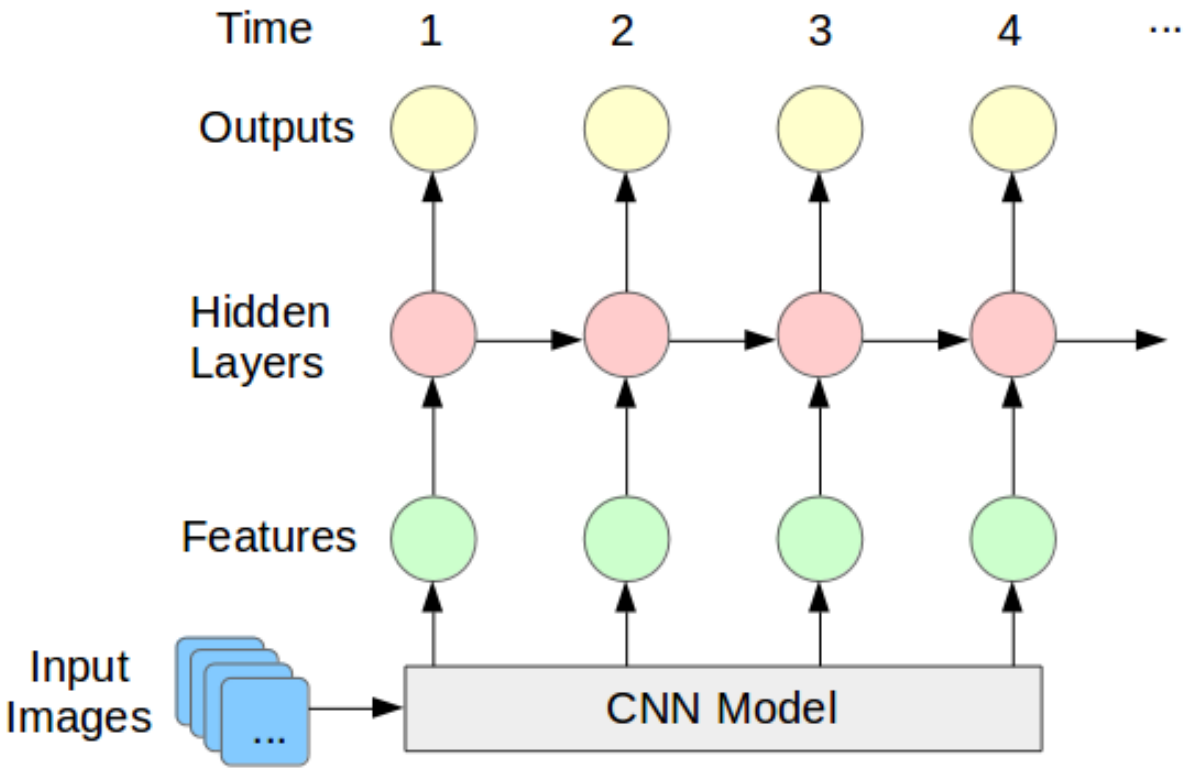
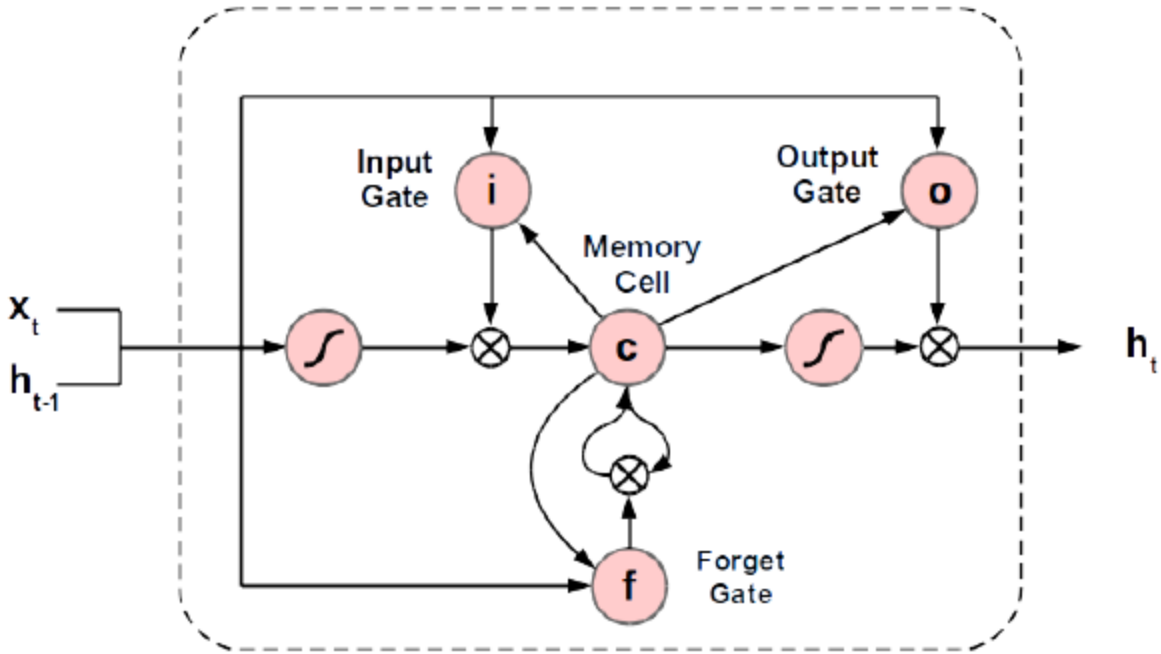


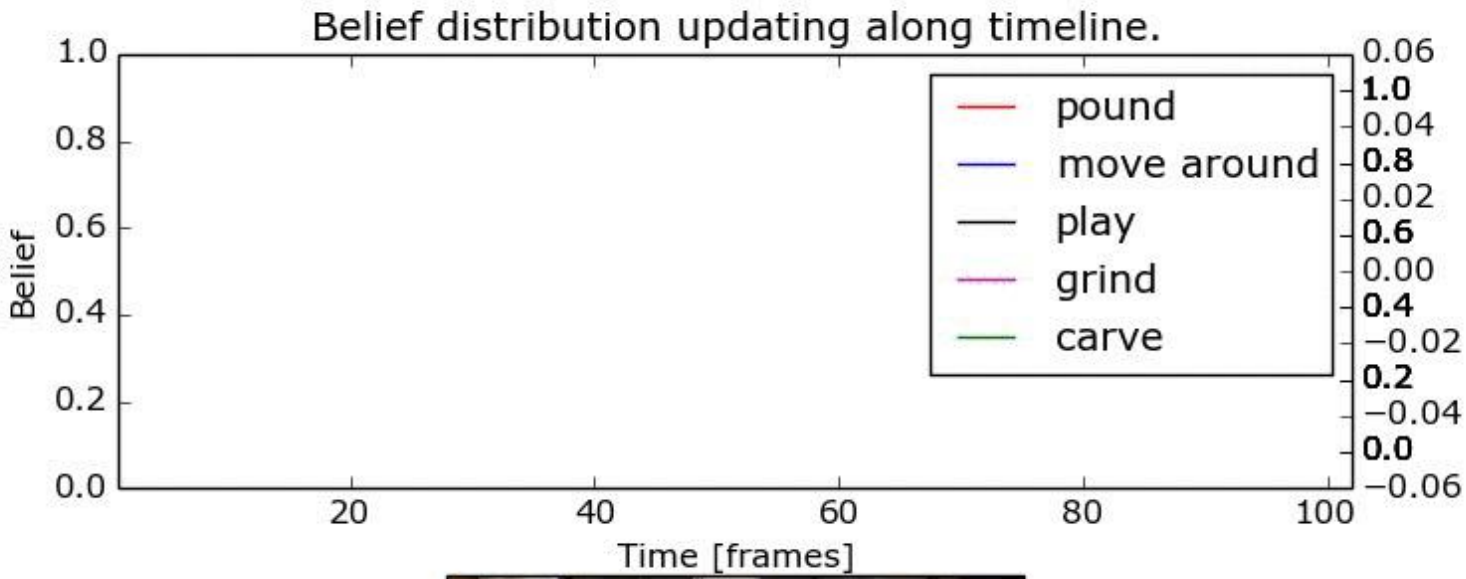
Image features from pretrained VGG model (4096 dimensions)

Diagram of a LSTM memory cell



These gating functions learn to control the portion of current input and previous memory the LSTM should take into consideration.

Action Prediction Results

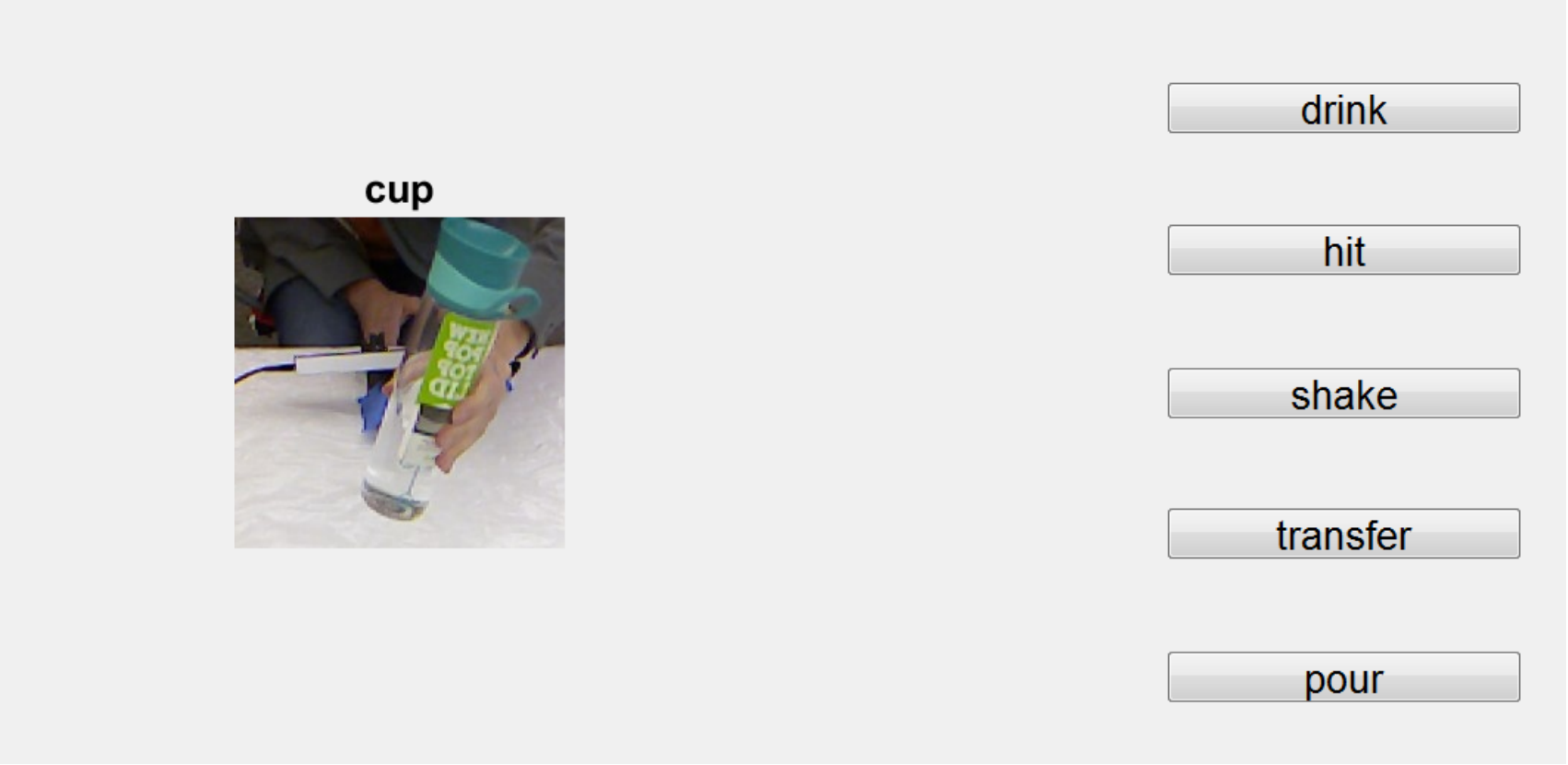


play

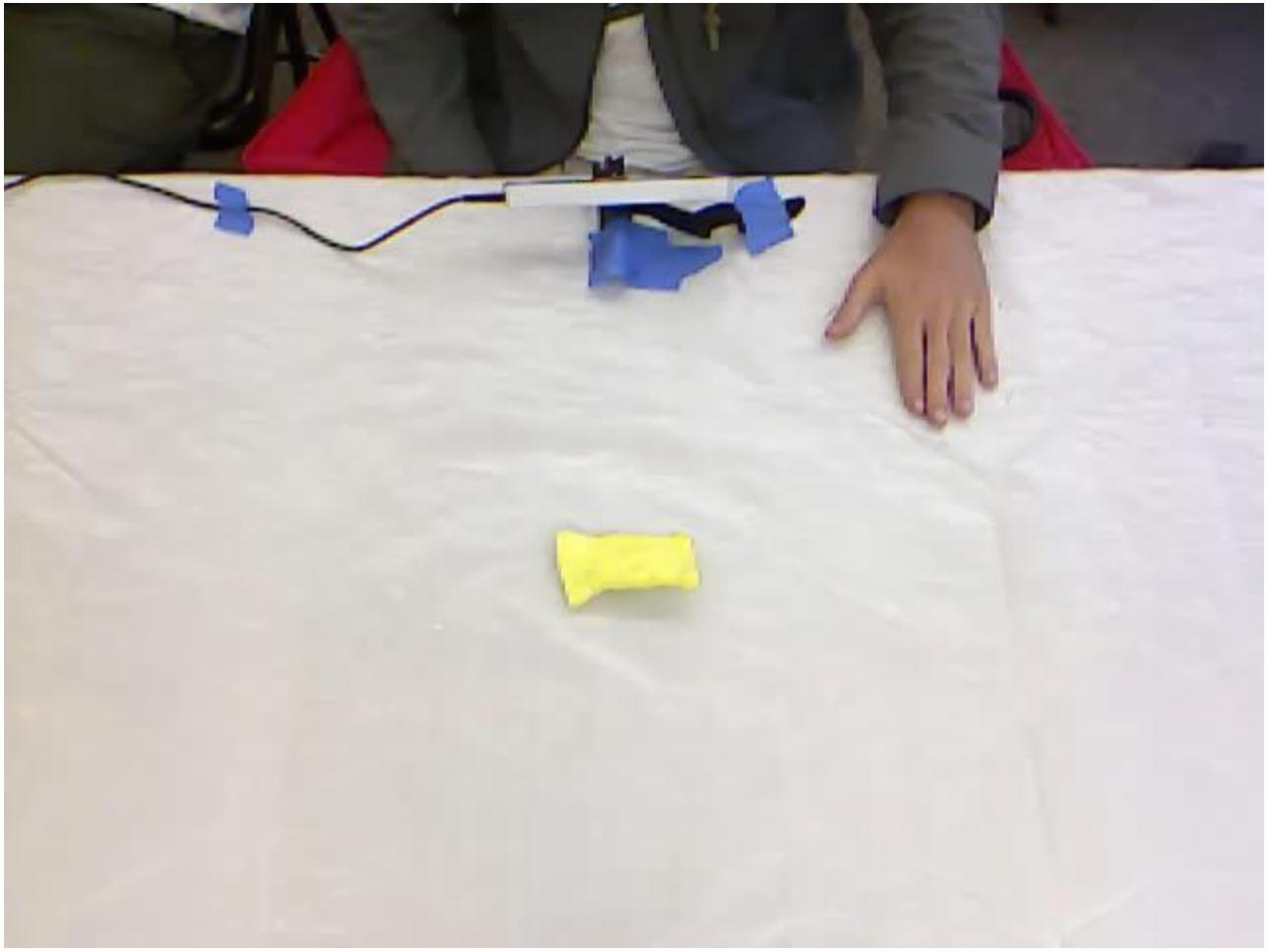
LSTM used as classifier

Psychophysics

to create ground truth data for evaluation

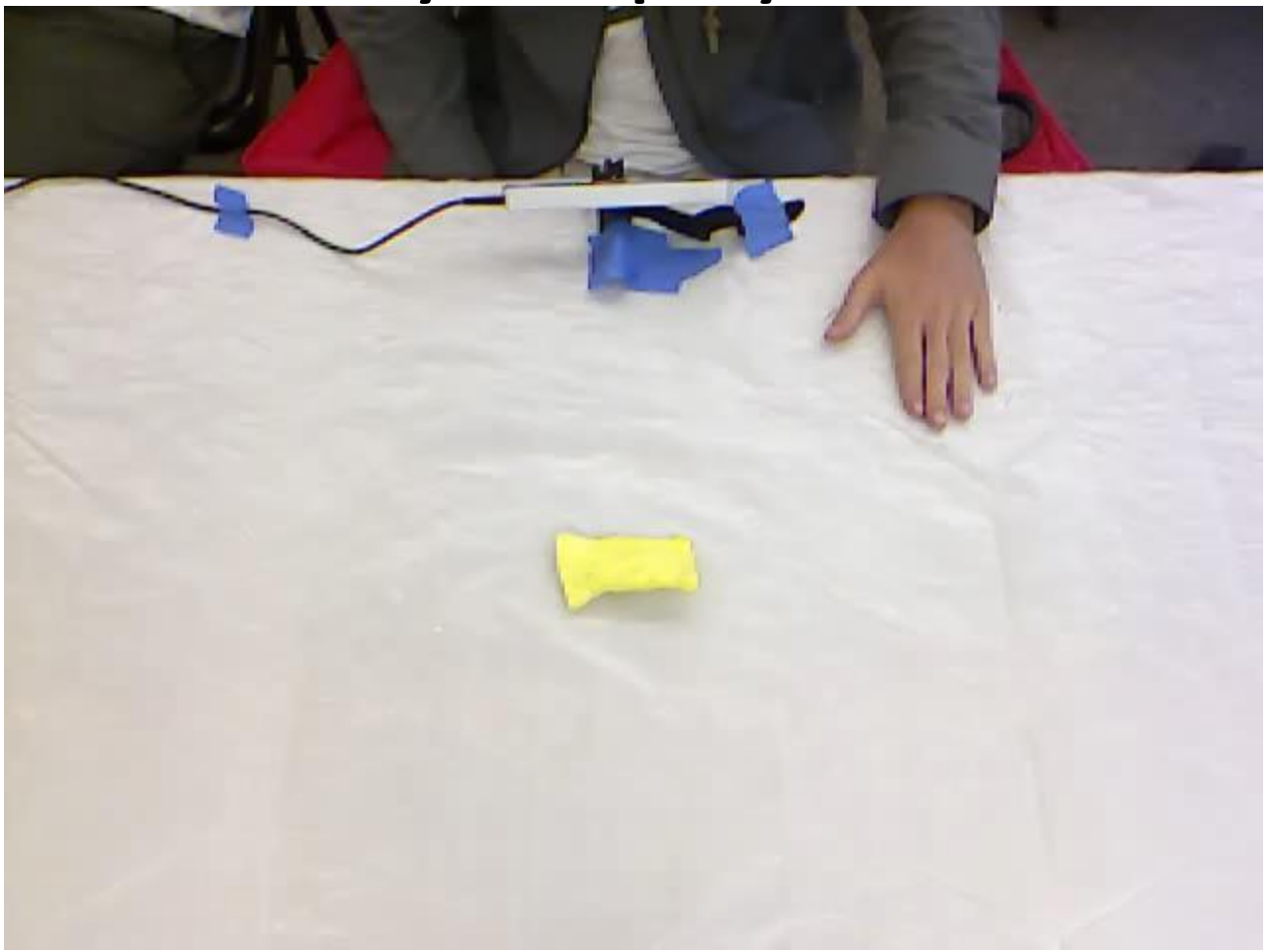


Psychophysics



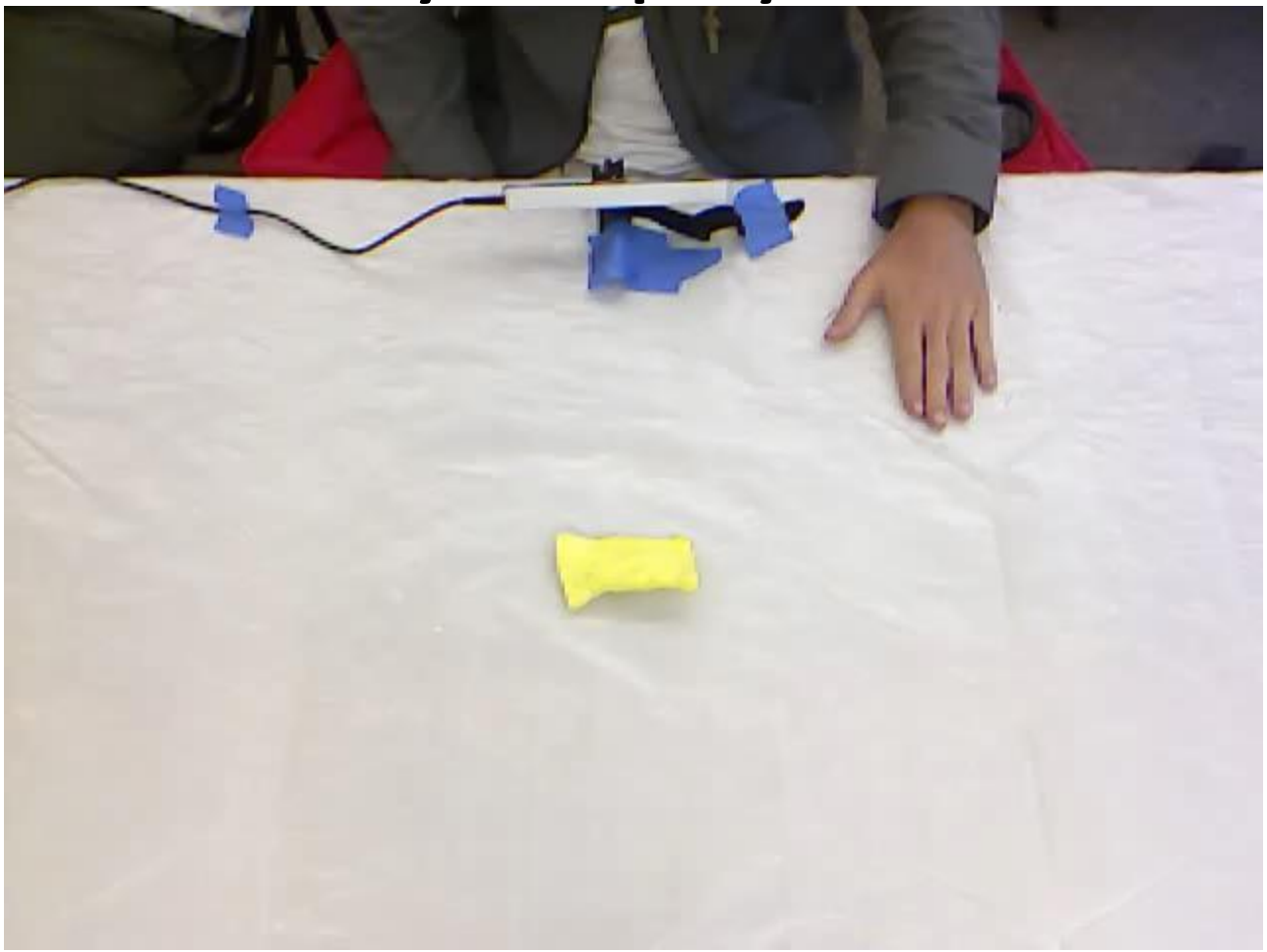
Extremely Hard:
Contact -10

Psychophysics



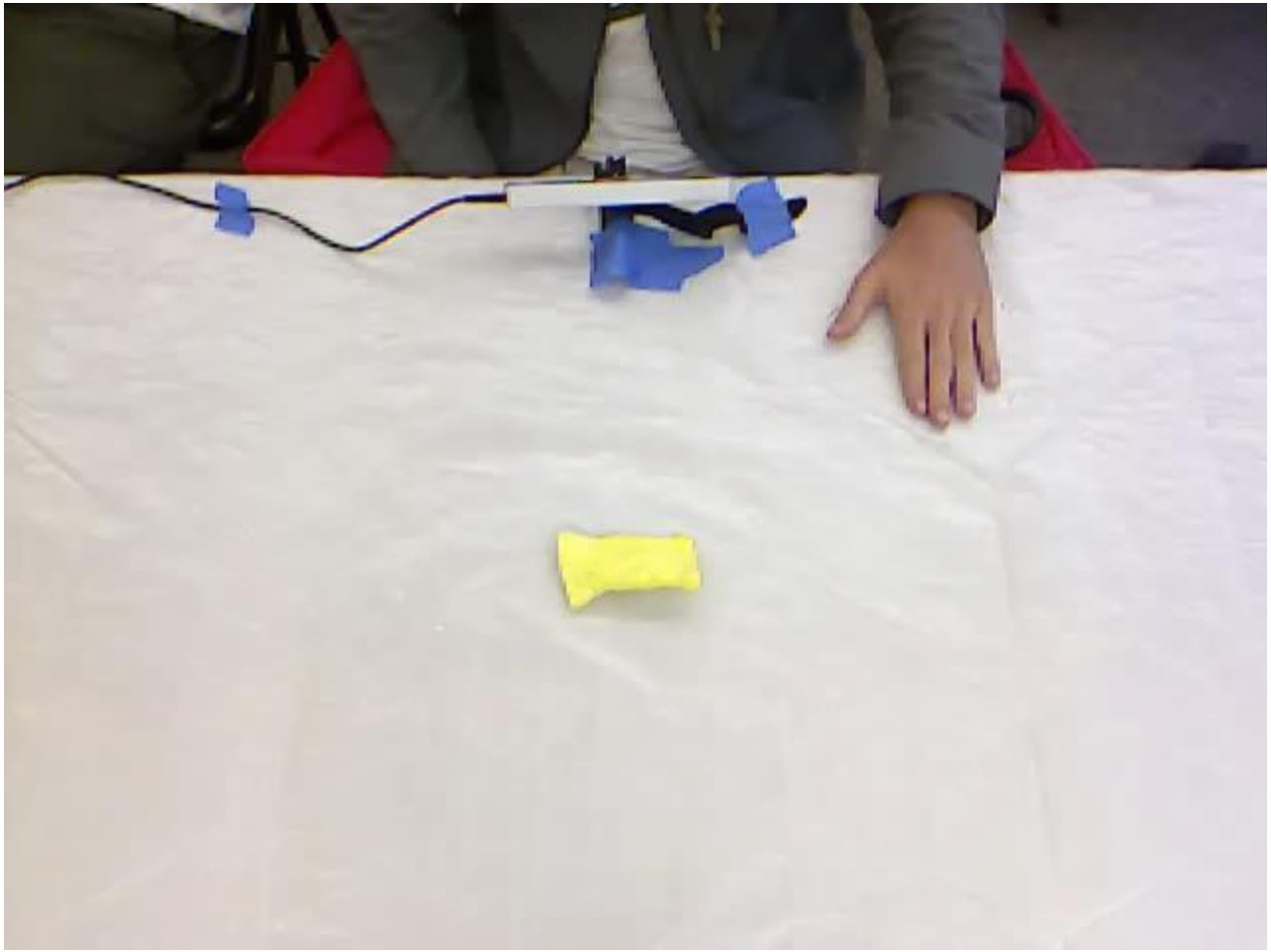
Hard: Contact

Psychophysics



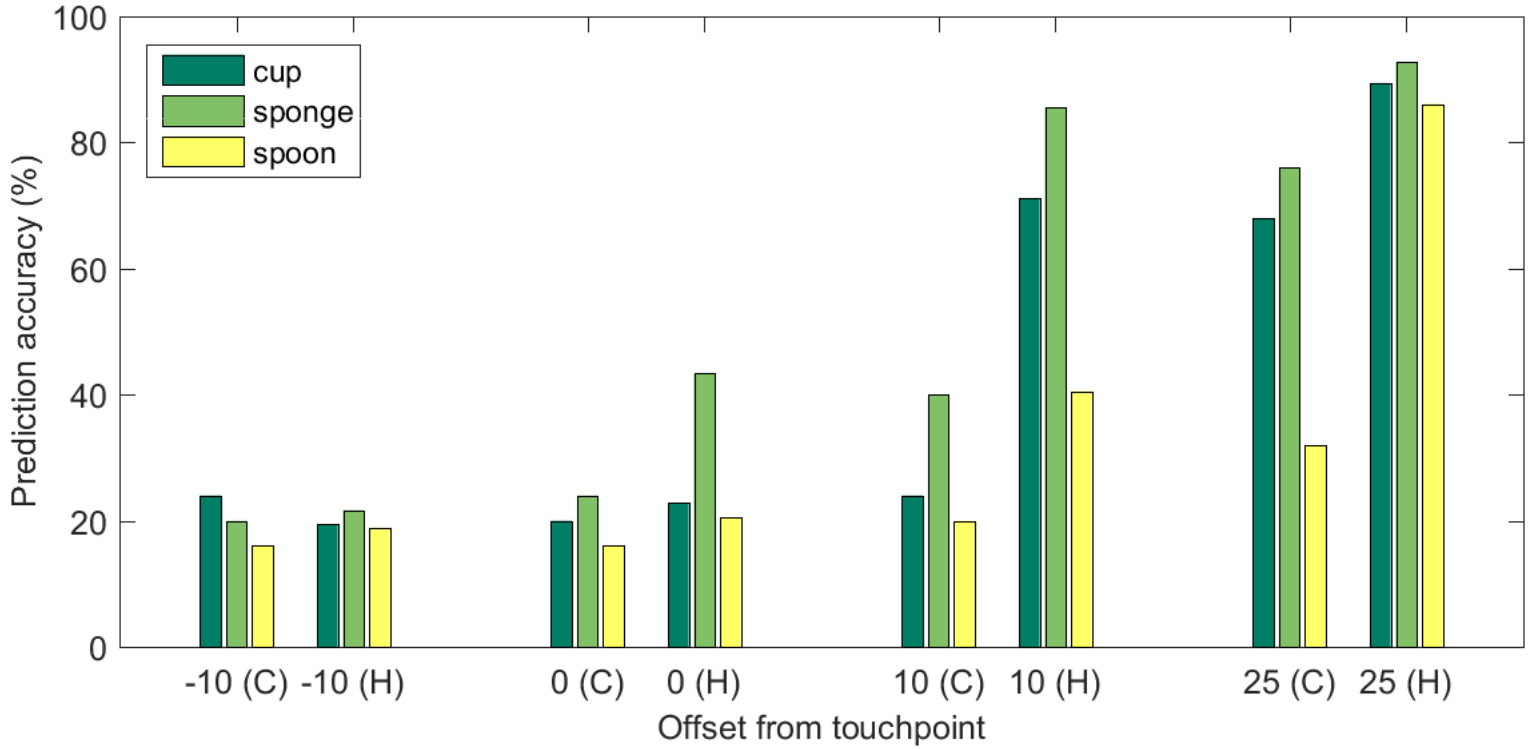
Easy: Contact +10

Psychophysics



Obvious: Contact +25

Prediction accuracy **H**uman vs **C**omputer

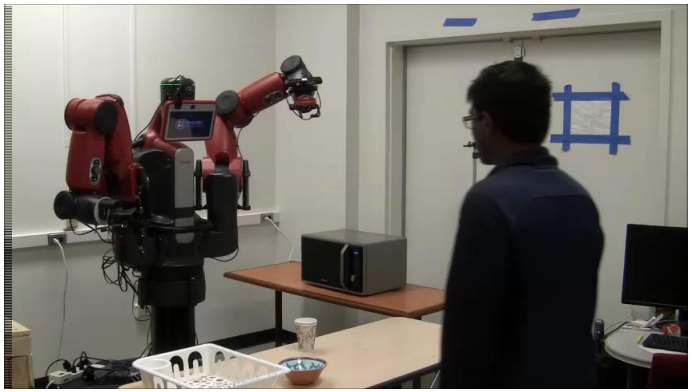


Why tactile force prediction

- Robot Learning

Classic Kinesthetic Teaching

instead



Learning from observing

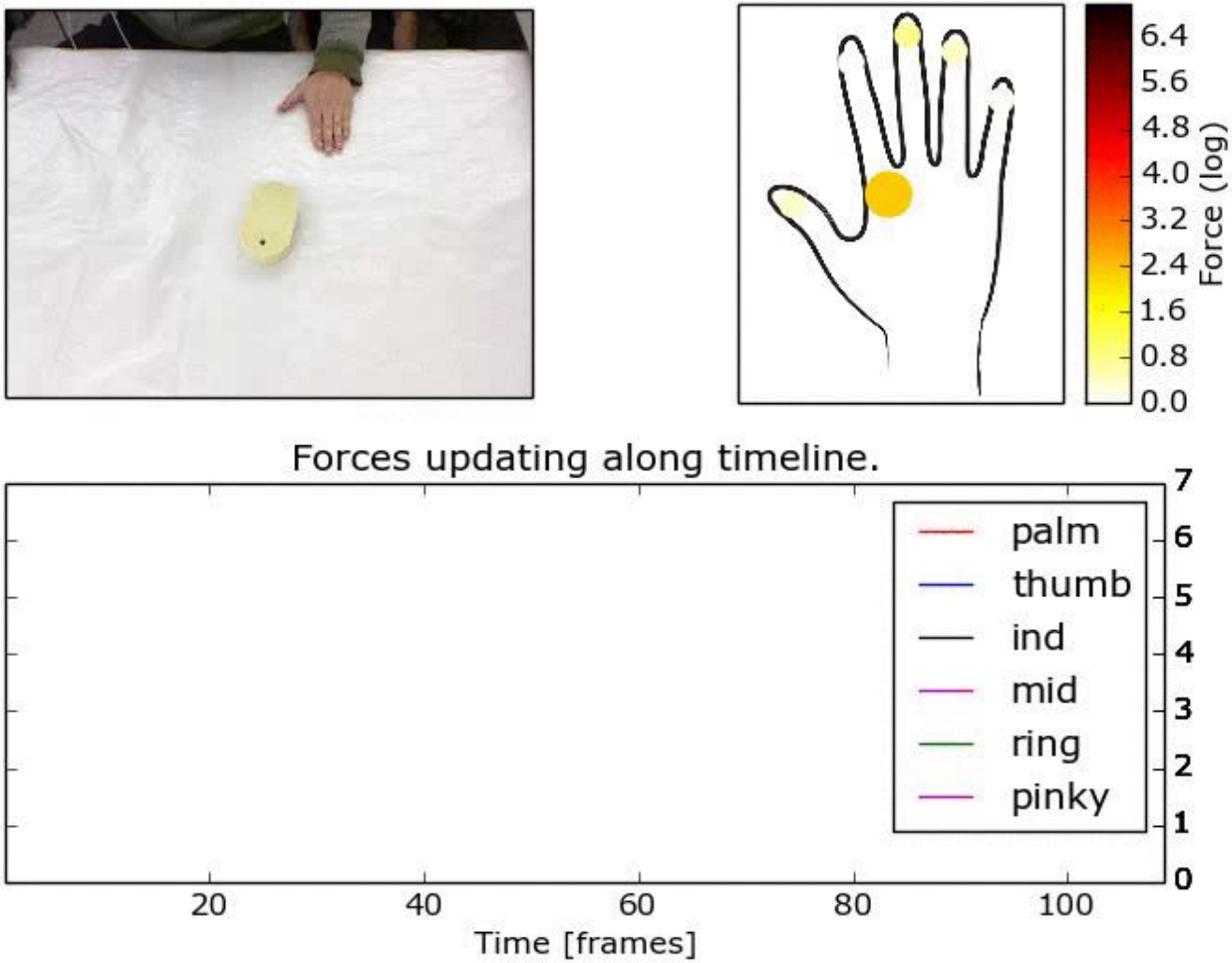
- Humans understand actions multimodal (Mirror neuron system)

Hand Force Estimation

- Data Collection
  - 5 subjects, 4 objects, 5 actions each, 500 samples
  - mirrored action with both hands

Force Estimation Illustration

LSTM used as regressor

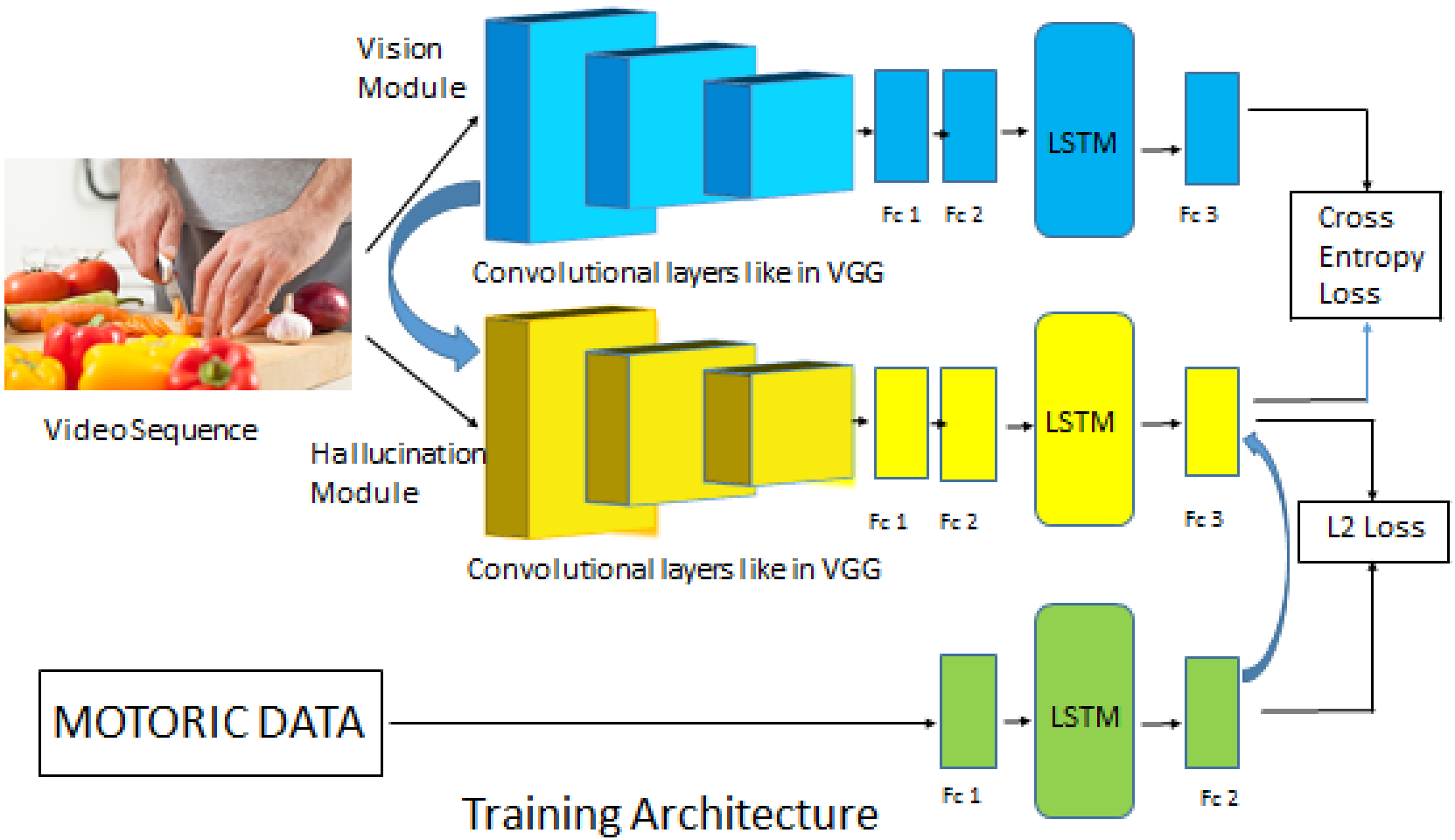


Better recognition (with force data)

Object	cup	stone	sponge	spoon	knife	Avg.
Vision	82.4	61.4	61.6	62.2	73.3	68.3
Vision + Force	88.2	75.1	59.1	57.5	72.7	70.5

Action prediction with images and predicted force (in %)
Training on Bimodal data; Predicting on Vision data

Learning with multimodal data



Telluride Neuromorphic Cognition Engineering Workshop

Telluride Neuromorphic Cognition Engineering Workshop